

①9 RÉPUBLIQUE FRANÇAISE  
INSTITUT NATIONAL  
DE LA PROPRIÉTÉ INDUSTRIELLE  
PARIS

①1 N° de publication :  
(à n'utiliser que pour les  
commandes de reproduction)

**2 729 236**

②1 N° d'enregistrement national : **95 00117**

⑤1 Int Cl<sup>6</sup> : G 05 B 19/402, 19/4097, G 06 T 1/40, B 25 J 9/10

⑫

## DEMANDE DE BREVET D'INVENTION

**A1**

②2 Date de dépôt : 06.01.95.

③0 Priorité :

④3 Date de la mise à disposition du public de la  
demande : 12.07.96 Bulletin 96/28.

⑤6 Liste des documents cités dans le rapport de  
recherche préliminaire : *Se reporter à la fin du  
présent fascicule.*

⑥0 Références à d'autres documents nationaux  
apparentés :

⑦1 Demandeur(s) : THOMSON BROADBAND SYSTEMS  
— FR.

⑦2 Inventeur(s) : VENAILLE CHRISTOPHE.

⑦3 Titulaire(s) :

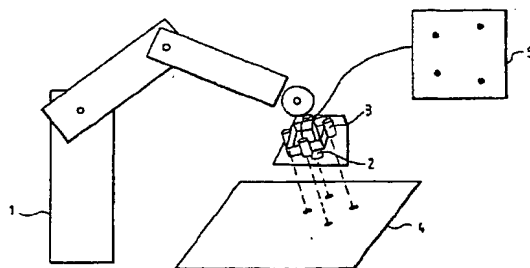
⑦4 Mandataire : THOMSON MULTIMEDIA.

### ⑤4 GUIDAGE DE ROBOT PAR ECLAIRAGE ACTIF.

⑤7 Le procédé de positionnement du robot (1) muni d'un capteur d'image (2) est réalisé à partir de la projection d'un motif lumineux sur la scène (4), projection (5) dont la forme est fonction de la position du robot.

L'asservissement se fait par comparaison de la forme de ce motif de l'image courante à la forme finale du motif de l'image de référence correspondant à la position recherchée, par l'intermédiaire d'un réseau de neurones recevant en entrées des informations relatives à la forme projetée du motif. Ces informations sont des coordonnées, des descripteurs de Fourier ou des moments géométriques.

Les applications concernent la robotique mobile telle que la conduite d'engins ou robotique industrielle.



FR 2 729 236 - A1



## GUIDAGE DE ROBOT PAR ECLAIRAGE ACTIF

La présente invention concerne un procédé de positionnement d'un robot dans l'espace à trois dimensions, ou 3D, et un robot évoluant  
5 dans un tel espace.

Le domaine technique dans lequel s'inscrit l'invention est celui de la "commande référencée capteur" d'un robot dont une application possible est la commande par asservissement du robot par rapport à un objet ou une scène donnée.

10 La "commande référencée capteur" est une approche nouvelle et originale pour la commande de robot : classiquement le déplacement d'un robot se fait par l'envoi de commandes qui indiquent, pour le mouvement désiré, les valeurs angulaires des rotations à appliquer séparément à chaque axe du robot (ces commandes sont généralement  
15 exprimées soit dans un repère cartésien, soit en coordonnées angulaires).

Pour amener avec une bonne précision l'organe terminal d'un robot à une position déterminée, une telle approche nécessite la connaissance géométrique exacte du modèle du robot et des calculs longs et complexes pour convertir dans l'espace cartésien, des mouvements  
20 définis dans l'espace articulaire. Par ailleurs, une telle approche ne prend absolument pas en compte les variations possibles de l'environnement et est donc dépourvue de toute adaptabilité.

L'approche proposée par la "commande référencée capteur" est la suivante : le but à atteindre par le robot est spécifié en termes de  
25 signaux devant être observés par des capteurs montés sur le robot, au lieu de mouvements à effectuer pour atteindre ce but.

Dans le cas d'une "commande référencée image", une caméra est fixée sur le robot et observe la scène qui se trouve devant le robot. L'image délivrée par la caméra varie donc selon la position du robot. La  
30 tâche d'asservissement du robot est spécifiée non pas en envoyant directement une commande de positionnement géométrique au robot, mais en donnant une image de référence de ce qui doit être observé par la caméra lorsque la position finale est atteinte.

Un module de "commande référencée capteur" calcule  
35 automatiquement les bons mouvements à appliquer au robot, à partir de la

différence observée entre une image réellement vue par la caméra et une image souhaitée dite de référence, en simplifiant ainsi considérablement la tâche demandée à l'opérateur.

L'avantage de la "commande référencée capteur" de robot est donc non seulement d'apporter une souplesse importante à l'utilisateur dans sa tâche de programmation du robot, mais aussi d'apporter enfin aux robots l'autonomie et l'adaptabilité qui leur font si cruellement défaut. En effet, les robots actuels ne peuvent travailler que sur des séquences répétitives apprises et pour lesquelles l'environnement est figé. En cas de changements même minimes de l'environnement, par exemple une petite erreur de placement d'un composant lors d'une tâche d'assemblage, les robots actuels ne sont pas capables de s'adapter à ces situations nouvelles. La "commande référencée capteur" offre cette capacité d'adaptation puisque dans ce cas le robot "voit ce qu'il fait" et est donc capable de s'asservir automatiquement par rapport à un objet pour lequel a été défini une image de référence.

Le problème général à résoudre est le calcul du déplacement d'un robot doté d'un capteur lorsqu'on connaît à tout instant le signal désiré, ou signal de référence, en position finale, ou position à atteindre, et le signal délivré par le capteur quand le robot est dans une position courante au voisinage de la position à atteindre.

Plusieurs approches ont déjà été proposées dans la littérature pour traiter le problème de la "commande référencée image" notamment :

- un article de F. Chaumette, P.Rives, B. Espiau intitulé "Positioning a robot with respect to an object tracking it and estimating its velocity by visual servoing", Proceedings of the 1991 IEEE International Conference on Robotics and Automation, pp.2248-2253, Sacramento, California, April 1991,
- un article de L.E. Weiss, C. Sanderson, C.P. Neuman intitulé: "Dynamic sensor-based control of robots with visual feedback", IEEE Journal of Robotics and Automation, Vol. RA-3, n° 5, pp.404-417, October 1987,
- un article de J.T. Feddema, O.R. Mitchell intitulé "Vision-guided servoing with feature-based trajectory generation", IEEE

Transactions on Robotics and Automation, Vol.5, n° 5, pp.691700, October 1989,

- un article de P. Rives, F. Chaumette, B. Espiau intitulé : "Visual servoing based on a task function approach", First International Symposium on Experimental Robotics, Montréal, Canada, June 1989, et
- un article de C. Samson, M. Le Borgne, B. Espiau intitulé : "Robot Control: The Task Function Approach", Oxford Science Publications, Oxford 1991.

Toutes ces approches suivent plus ou moins le même principe décrit ci-dessous.

Tout d'abord, un certain nombre de paramètres sont déterminés hors asservissement, par les étapes telles que :

- la définition de l'image de référence, consistant à choisir préalablement une image de référence de l'objet ou de la scène vue par le robot qui constituera la cible à atteindre par le robot, l'image à obtenir par déplacement du robot.
- la sélection des caractéristiques "image" consistant, dans l'image de référence à isoler un nombre fini de caractéristiques (point, lignes remarquables, etc...),
- la modélisation 2D/3D consistant à identifier pour chacune de ces caractéristiques les coordonnées 2D dans l'image et les informations 3D associées dans la scène sur l'objet ou dans la scène observée par la caméra du robot,
- la calibration "robot/caméra" pour la détermination du modèle géométrique exact du robot et la position relative exacte de la caméra par rapport au repère géométrique associé au robot, et
- le calcul de la matrice jacobienne de la tâche d'asservissement obtenu à partir de relations géométriques entre les points caractéristiques 3D de la scène, le robot et le capteur.

Le calcul théorique du déplacement  $df$ , dans l'image, des coordonnées des points caractéristiques sélectionnés en fonction du déplacement du robot  $dX$  peut être alors effectué à partir de la matrice jacobienne de la tâche d'asservissement, par la formule suivante :

$$df = J_{\text{tâche}} \cdot dX \quad (1)$$

En fait, c'est donc la détermination de l'inverse de cette matrice tâche qui est intéressante, de façon à retrouver la commande à appliquer au robot lorsque la déviation image  $d_f$  est connue.

Une fois ces données acquises, le processus d'asservissement visuel, ou "commande référencée capteur" se déroule suivant un processus itératif.

Le robot reçoit un ordre de déplacement en fonction des déviations images. Une nouvelle image est alors acquise et à nouveau comparée à l'image de référence. Le robot est ainsi déplacé jusqu'à convergence vers la position désirée. A chaque itération, les opérations suivantes doivent être réalisées :

- une acquisition et un traitement de l'image courante,
- une extraction des caractéristiques "images" et un appariement consistant, dans l'image acquise, à détecter les points caractéristiques et à les mettre en correspondance avec les points sélectionnés dans l'image de référence, et

- un calcul du mouvement inverse qui à partir des différences entre les deux images, détermine la commande à appliquer au robot par inversion de la matrice jacobienne de la tâche d'asservissement.

Les inconvénients liés à ce type d'approche limitent la portée du concept de "commande référencée capteur" par le manque de fiabilité de la méthode, notamment du fait d'une grande sensibilité des phases de traitement de l'image et de l'appariement, par la lourdeur de mise en oeuvre (nécessité de modélisation de l'ensemble scène, robot et caméra), et par le faible champ applicatif possible (nécessité de voir des éléments "facilement reconnaissables" par un algorithme dans l'image observée).

Ces inconvénients sont énumérés ci-dessous :

Les étapes de modélisation 2D/3D et de calibration "robot/caméra" exigent une bonne connaissance des positions géométriques relatives des différents constituants que sont le robot, la caméra, et les caractéristiques 3D sur lesquelles l'asservissement est réalisé. Cela ne peut être obtenu que par des moyens demandant un effort important de modélisations et de calibrations qui sont des tâches réputées fastidieuses et imprécises.

De plus, l'étape du calcul théorique du jacobien de la tâche limite le type de données caractéristiques pouvant être utilisées à des primitives géométriques simples telles que des points, segments de droites, cercles, ellipses etc..., et qui sont définies localement. En aucun cas elles ne peuvent être étendues à des descripteurs globaux de l'image dont la détermination est pourtant plus robuste.

Les étapes d'acquisition et traitement de l'image courante, d'extraction des caractéristiques "images" et d'appariement sont les plus délicates et sont très sensibles au bruit : le principe de la méthode impose en effet la possibilité, et cela quelles que soient les conditions d'éclairage, la position du robot par rapport à l'objet, etc..., d'extraire dans l'image vue par le robot les éléments caractéristiques recherchés, et de les apparier à ceux de l'image de référence. Toute erreur à ce niveau entraîne généralement un échec de la méthode.

En conclusion, la recherche d'une relation mathématique entre les mouvements du robot et les déviations observées dans l'image, par rapport à un signal de référence, est un problème non-linéaire difficile qui exige un important effort de modélisation et qui est généralement très sensible aux algorithmes de traitement d'images utilisés pour l'extraction de primitives.

En complément de toutes ces approches, dirons nous classiques, apparaissent de nouveaux procédés basés sur l'exploitation de réseaux de neurones qui permettent de résoudre un certain nombre des problèmes soulevés.

Les informations recueillies par le calculateur et provenant du ou des capteurs montés sur le robot sont traitées en temps réel en alimentant un réseau de neurones dont les sorties commandent le déplacement du robot. Ce réseau aura préalablement été entraîné, cette phase d'apprentissage classique permettant de définir ses poids synaptiques à partir d'exemples correspondant généralement à des positions du robot proches de sa position de consigne.

Mais, lorsque ce capteur est un capteur d'images, des inconvénients propres à ce procédé subsistent :

Il y a d'abord la contrainte d'éclairement de la scène, éclairement qui doit être supérieur à une certaine valeur et qui doit d'autre part se rapprocher de l'éclairement utilisé lors de la phase d'apprentissage.

Les caractéristiques propres de la scène peuvent rendre le  
5 procédé basé sur une comparaison d'images moins robuste; faire de la superposition d'image est plus délicat pour des images uniformes, sans grand contraste.

Le nombre d'exemples à apprendre par le réseau de neurones peut être relativement important pour des scènes, des objets complexes,  
10 c'est à dire à forte variation d'apparence en fonction de l'angle de vue; il s'agit par exemple d'objets ayant un nombre très élevé de facettes. Dans ce cas, la recherche de l'image identique est plus longue, le traitement plus complexe et le temps de réponse du robot plus long.

Le choix d'un vecteur  $f$ , permettant de définir la position de  
15 l'image par ses coordonnées de points caractéristiques de l'image, doit être renouvelé pour chaque scène, c'est à dire chaque image de consigne, lorsque ces scènes sont très différentes et ne comportent pas ces mêmes points caractéristiques. Les dimensions de vecteur peuvent être importantes si l'on désire obtenir une bonne précision, générant des  
20 calculs longs et coûteux.

Le but de l'invention est de pallier tous ces inconvénients. A cet effet, l'invention a pour objet un procédé de positionnement d'un robot évoluant dans un espace en 3 dimensions dont la scène est visualisée par un capteur d'image solidaire du robot, caractérisé en ce qu'un motif  
25 lumineux est projeté sur la scène dans le champ du capteur, en ce que la forme projetée du motif est fonction de la position relative du robot par rapport à la scène et en ce que le positionnement est réalisé par comparaison, en entier ou en partie, de la forme de ce motif de l'image courante à la forme finale du motif de l'image de référence correspondant à  
30 la position recherchée, par l'intermédiaire d'un réseau de neurones recevant en entrées, des informations relatives à la forme projetée du motif acquise par le capteur.

Le procédé de positionnement ainsi caractérisé concerne l'exploitation d'images d'une scène généralement figée mais pouvant  
35 évoluer d'un positionnement du robot à l'autre et dont le point de vue qui

est la position du capteur d'image, varie. La source lumineuse, solidaire du capteur, projette un motif sur la scène. La variation de la forme projetée, lors du déplacement du robot, permet d'asservir ce dernier à une forme particulière du motif mémorisée dans une image de référence  
5 correspondant à la position de consigne.

Les informations exploitées dans l'image sont généralement les paramètres d'un vecteur  $f$ , caractérisant la position de cette image, c'est à dire de la caméra, dans la scène 3D. Le nombre de paramètres, par exemple des descripteurs de Fourier, peut être important lors d'un  
10 traitement d'image classique. L'invention permet ici de limiter le nombre de paramètres traités, fonction du motif lumineux choisi et de la précision souhaitée et donc faciliter le traitement d'informations et la rapidité de calcul. Les algorithmes de traitement sont également simplifiés. En effet, l'exploitation classique de paramètres intrinsèques à la scène et dont le  
15 choix est toujours fastidieux et difficile est ici supprimé du fait que l'on intervient directement dans la scène pour générer les paramètres définissant le vecteur  $f$ . Ceux-ci sont indépendants de la scène choisie puisque fonction du motif projeté.

Le procédé selon l'invention est également plus robuste, du fait  
20 de la capacité d'extraction des informations liées à l'éclairage actif :

- soit l'éclairage de la scène peut être contrôlé et la séparation des motifs lumineux est assurée par réduction de l'éclairage ambiant de la scène.
- soit l'éclairage n'est pas maîtrisable et l'utilisation de lasers ou  
25 de filtres optiques associés à la longueur d'onde de l'émission lumineuse ou même de caméra infra-rouge assurent l'extraction de l'information utile. Il en découle, lorsque ces 2 dernières solutions sont choisies, que l'exploitation du robot devient indépendante des conditions d'éclairage ambiant et c'est là un autre avantage important.

30 Une description de différents exemples de réalisation est faite ci-après à l'aide des figures qui représentent :

- figure 1, un robot avec caméra et éclairage actif,
- figure 2, un support mécanique des sources lumineuses,
- figure 3, les images correspondant à la position courante et la  
35 position de consigne du robot.



- figure 4, une chaîne algorithmique de traitement des images exploitant les barycentres,
- figure 5, une deuxième chaîne algorithmique exploitant les descripteurs de Fourier,
- 5       - figure 6, un robot avec caméra et projection d'un motif lumineux en forme de grille,
- figure 7, une troisième chaîne algorithmique exploitant les moments géométriques,
- 10       - figure 8, une quatrième chaîne algorithmique exploitant la corrélation.

La figure 1 représente un robot 1 muni d'une caméra 2 en bout d'articulation. Cette caméra se trouve au centre d'une structure de forme carrée ou cadre supportant 4 lasers 3, ceux-ci étant fixés au milieu de chaque côté du cadre. Les sources lumineuses définissent ainsi un motif  
15 lumineux correspondant à quatre taches aux sommets d'un carré, lorsque la surface de projection est perpendiculaire aux faisceaux laser parallèles. Ce motif projeté sur la scène 4 est déformé par la surface sur laquelle il est projeté pour fournir une image 5.

Le système, réalisant une phase d'apprentissage, ne nécessite  
20 pas un parallélisme des faisceaux laser, le motif constitué des 4 points pouvant être quelconque. Les calculs sont cependant facilités si tel est le cas, c'est à dire si le motif est facilement identifiable. Le positionnement des lasers doit également tenir compte de l'axe optique et de la focale de l'objectif de la caméra, de manière à ce que les taches soient dans le  
25 champ de la caméra dans la phase de mise en application du procédé, c'est à dire au voisinage de la position de consigne du robot. Cette focale est d'ailleurs liée à la précision du dispositif.

Pour ce faire, la figure 2 détaille un exemple de cadre support de moyens de fixation 7 des lasers 8, fixation à 2 degrés de liberté pour  
30 chacun d'entre eux, rotation autour de l'axe 9 et autour de l'axe 10, permettant de régler leur positionnement.

La figure 3 représente le robot 1 a sa position courante 1-1, la caméra solidaire fournissant une image courante 5-1, puis a sa position de consigne 1-2, l'image obtenue 5-2 étant alors l'image de référence.

La figure 4 représente l'algorithme de traitement de l'image captée par la caméra fixée au robot et constituée par la projection de ce type de motif lumineux.

L'image est, dans une première étape 11, captée par la caméra  
5 par exemple de type CCD et transformée en signaux électriques pour fournir une image de gris. C'est par exemple une image de 512x512 pixels donc constituée de 262 144 informations.

Une deuxième étape effectue un seuillage 12 sur la luminance des pixels pour ne retenir que ceux supérieurs à une valeur prédéterminée.  
10 L'image binaire obtenue est traitée, dans l'étape 13, par recherche des composantes connexes; cette recherche consiste à attribuer à une zone les pixels connexes, cette zone connexe représentant alors la surface d'une tache lumineuse dans l'image formée par la projection d'un faisceau laser sur la scène.

15 Une quatrième étape 14 réalise le calcul du barycentre des pixels correspondant à la zone précédemment déterminée. Lors de cette étape est donc calculé le "centre" de chacune des taches lumineuses, c'est à dire des coordonnées u et v de leur barycentre dans l'image fournie. Ce centre de la tache peut ne pas correspondre à un pixel.

20 Ce calcul de barycentre est donné à titre d'exemple. Il est en effet tout aussi possible, lors de cette étape et pour déterminer un centre de tache, de réaliser une convolution de l'image avec une tache parfaitement ronde, en prenant comme centre celui de cette tache lorsque son emplacement correspond au maximum de corrélation avec la tache de  
25 l'image captée.

L'étape 15 effectue un ordonnancement de ces centres, c'est à dire une affectation à chacun des lasers, par exemple en les associant, lors du traitement, à chacun des quadrants de l'image pour constituer un vecteur f de dimension 8, dont les coordonnées sont les coordonnées u et  
30 v de chacun des centres des tâches selon l'ordre défini. Ce vecteur caractérise l'image donc le point de vue pour une scène fixe, donc la position de la caméra.

$$f = (u_1, v_1, u_2, v_2, u_3, v_3, u_4, v_4)$$

Les étapes 16, 17, 18, 19 et 20 sont en tout point identiques aux  
35 étapes 11 à 15 mais concernent l'image de référence et non pas l'image

courante. Cette image de référence correspond à l'image de la scène visualisée par la caméra lorsque le robot est à sa position de consigne, position recherchée. Le vecteur  $f_{ref}$  défini à partir de cette image de référence est donc de même dimension et ordonnancé de la même manière que  $f$  et est comparé en permanence au vecteur  $f$  lors de l'étape 21. Cette étape effectue donc une différence terme à terme entre ces 2 vecteurs pour fournir, à l'étape 22, un vecteur  $df$  représentant un signal d'erreur lié à l'écart de position du robot entre sa position réelle et la position finale désirée définie par l'image de référence. Ce signal d'erreur alimente ensuite un réseau de neurones, chaque entrée du réseau recevant un terme du vecteur différence. La sortie de ce réseau délivre alors un vecteur déplacement au robot, comme il sera vu plus loin, le réseau ayant initialement subi une phase d'apprentissage dont le principe est connu. La dimension des vecteurs  $f$  peut être différente de 8. Il est ainsi possible de n'exploiter que 3 taches et donc 3 lasers ou un nombre de taches supérieur à 4.

L'exploitation de ce procédé nécessite évidemment que les points caractéristiques de l'image créés artificiellement, que constituent les taches lumineuses formées par les lasers, soient présents dans la scène et visualisés par la caméra et que le motif généré par ces points soit identifiable. Ainsi, lorsque les objets sur lesquels sont projetées ces taches sont à forte discontinuité ou que, par exemple, un faible mouvement de la caméra entraîne une modification importante du motif et c'est le cas lorsque une surface de projection évolue au voisinage d'un plan parallèle au faisceau laser, la corrélation avec le motif de référence peut être délicate. Il est alors préférable d'utiliser un motif lumineux plus complexe, par exemple un motif à grille ou à ligne continue. Le principe du choix est de conserver un motif se déformant de manière continue lors du déplacement de la source lumineuse afin de faciliter la correspondance ou corrélation avec l'image de référence. Il ne faut pas que l'empreinte du motif sur la scène se déforme "d'un seul coup" lors d'un petit déplacement du robot. Ceci afin de toujours pouvoir "caractériser" l'image, c'est à dire reconnaître les points caractéristiques pour reconstituer le vecteur  $f$ . Les informations intrinsèques portées par l'empreinte, par exemple la position des lignes et des colonnes dans le cas d'un motif lumineux en forme de

grille, doivent être d'autant plus nombreuses que la surface sur laquelle est projeté le motif lumineux est discontinue, afin de faciliter la corrélation.

La réalisation de ces motifs lumineux peut se faire à partir de mires de projection, source lumineuse autocollimatée associée à un écran  
5 ajouré selon le motif à projeter sur la scène, cet "ajourage" pouvant être réalisé électroniquement par l'emploi d'éléments à cristaux liquides. Il peut s'agir également d'un balayage particulier d'un ou plusieurs faisceaux laser ou l'association de lentilles semi-cylindriques de dispersion à ces lasers. D'une manière générale, tout dispositif réalisé à partir d'une lumière  
10 structurée, c'est à dire avec une source active d'éclairage permettant de faire une empreinte sur la scène, convient

L'exploitation du procédé précédemment décrit nécessite d'autre part une mise en correspondance de points du motif, relativement facile lorsqu'il s'agit d'un motif constitué de quelques points mais source d'erreurs  
15 lorsque ce motif devient plus complexe.

Une solution à ces problèmes est l'utilisation de descripteurs globaux, relatifs à une forme projetée fermée ou quelconque. L'image, en fait le motif projeté, est caractérisée de manière globale, sans qu'il soit nécessaire d'affecter les paramètres aux sources lumineuses, de les  
20 ordonnancer. Ces descripteurs globaux permettent, quel que soit leur nombre, fixé, de toujours faire correspondre des formes quelconques.

Si le motif projeté est une forme fermée, par exemple un cercle ou un carré, les descripteurs globaux les mieux adaptés sont les descripteurs de Fourier et le procédé, en regard de l'algorithme représenté  
25 en figure 5, est décrit ci-après.

L'image courante captée par la caméra à l'étape 23 est seuillée à l'étape 24. L'étape 25 effectue une extraction des contours, de manière connue. C'est par exemple une extraction de régions avec sélection des pixels les plus à l'extérieur, intérieur ou au centre de la zone représentant  
30 le motif fermé. Un calcul des descripteurs de Fourier est ensuite effectué à l'étape 26 à partir de cette nouvelle image. Il s'agit, à partir des coordonnées d'un nombre N de points  $U_m$  sélectionnés sur le contour et équidistants, c'est à dire du contour échantillonné, de calculer N descripteurs de Fourier définis par :

$$D_n = \frac{1}{N} \sum_{m=0}^{N-1} U_m e^{-j2\pi \frac{nm}{N}}$$

avec  $U_m = x_m + j y_m$ ,  $x_m$  et  $y_m$  étant les coordonnées du point  $U_m$  et avec  $n$  variant de 0 à  $N-1$ .

- 5 On travaille donc dans le plan complexe. Sur les  $N$  paramètres calculés, on conserve  $2K+1$  paramètres, soit  $n$  variant de 0 à  $K$  et de  $N-K$  à  $N-1$ ,  $K$  étant un entier. Le choix de  $K$  correspond tout simplement à un filtrage. Diminuer  $K$  revient à réaliser un filtrage passe-bas, c'est à dire à "lisser" la forme décrite lorsque l'on reconstitue, à partir des descripteurs
- 10 sélectionnés, cette forme dans le plan réel. Ainsi, à l'extrême, le choix du descripteur de plus basse fréquence correspond au barycentre de la forme fermée, des deux descripteurs de plus basse fréquence à un cercle et ainsi de suite en se rapprochant de plus en plus de la forme fermée initiale.

- L'intérêt d'exploiter des descripteurs de Fourier se trouve donc
- 15 dans le fait qu'ils permettent une représentation vectorielle de taille fixe, bien adaptée au nombre d'informations en entrée du réseau de neurones qui doit être constant, et cela en conservant la cohérence dans les paramètres à comparer entre l'image de référence et l'image courante. En d'autre termes, quel que soit le nombre choisi, les formes définies par ces
- 20 descripteurs seront toujours comparables, le nombre étant lié à la précision. L'étape 27 sélectionne les descripteurs basse fréquence, le nombre correspondant à celui des entrées du réseau de neurone. Ce nombre d'entrées sera donc défini, à l'origine, en fonction de la précision souhaitée, puisque lié à un filtrage basse fréquence, donc à une perte
- 25 d'informations. L'étape 28 réalise, d'une manière connue, une normalisation des descripteurs de Fourier : le décalage du point de départ sur le contour agissant sur les descripteurs de Fourier, cette opération de normalisation permet de rendre les descripteurs normalisés invariants par rapport au point de départ choisi. L'avantage est évident : si par exemple
- 30 le traitement de l'image se fait en prenant, comme point de départ, le point le plus en haut à gauche de l'image de référence, ce point n'est pas forcément le même que celui de l'image courante, pour laquelle l'orientation de la caméra peut être différente. La comparaison des descripteurs simples nécessiterait donc une mise en correspondance.

Grâce aux descripteurs de Fourier normalisés, cette mise en correspondance, source de calculs et d'erreurs, n'est pas nécessaire. L'invariance par rapport au point de départ rend l'ordonnement inutile.

Les opérations décrites aux étapes 23 à 28 et effectuées sur l'image courante sont réalisées de la même manière sur l'image de référence, aux étapes 29 à 34. Mais ici, le calcul de  $f_{ref}$  n'est réalisé qu'une fois sur l'image de référence et non pas pour chaque image courante, le résultat étant mémorisé. Le comparateur 35 effectue la comparaison des descripteurs de Fourier normalisés de l'image de référence, constituant le vecteur  $f_{ref}$  mémorisé, avec ceux de l'image courante constituant le vecteur  $f$ . Le vecteur différence  $df$ , obtenu à l'étape 36 alimente l'entrée du réseau de neurones par ses composantes.

Lorsque les motifs projetés ne sont pas des formes simples fermées mais des formes plus complexes, un algorithme de traitement basé sur l'exploitation des moments géométriques de l'image est utilisé. Cette plus grande complexité permet de mieux caractériser l'image de par le fait que le motif peut être uniformément réparti sur la scène, et est donc particulièrement nécessaire pour des scènes à structure complexe, irrégulière. Il s'agit par exemple d'une grille formée de lignes et de colonnes, comme représentée à la figure 6. La figure 7 schématise l'algorithme traitant ce type de motif lumineux.

Après prise en compte de l'image courante, c'est à dire traduction en signaux vidéo de l'image de la scène et du motif projeté sur elle, étape 37, ces signaux subissent un prétraitement à l'étape 38. Ce prétraitement permet de réduire le nombre d'informations à traiter, donc d'accélérer le traitement dans son ensemble, par exemple en effectuant un seuillage de l'image comme explicité précédemment. Il peut également s'agir d'un calcul de gradient de luminosité de l'image pour ne conserver que les pixels correspondant à un gradient supérieur à une certaine valeur. Ce prétraitement permet aussi de s'affranchir du bruit, qu'il soit vidéo ou dû à l'éclairage, par exemple les réflexions parasites. L'étape 39 réalise le calcul des moments géométriques de l'image courante observée, moments géométriques  $m_{p,q}$  d'ordre  $p+q$  définis par la relation :

$$m_{p,q} = \sum_{x=0}^1 \sum_{y=0}^c x^p \cdot y^q f(x,y) \cdot \frac{1}{l.c}$$

$f(x,y)$  est la luminance du point de coordonnées  $x$  et  $y$ ,  $l$  et  $c$  sont le nombre de lignes et colonnes de l'image,  $p$  et  $q$  sont des entiers.

Par exemple, le calcul des moments géométriques d'ordre 2 permet de définir l'axe principal d'inertie de l'image considérée. Il y a une infinité de moments géométriques et, comme pour les descripteurs de Fourier, plus le nombre retenu est important, plus précise est l'image reconstituée à partir de ces moments géométriques. Ici aussi, le nombre correspond à celui des entrées du réseau de neurones. Le choix est effectué en sélectionnant (donc en calculant) les moments d'inertie d'ordre faible. Ceci permet d'éviter les amplifications importantes du bruit, c'est à dire des variations importantes des moments géométriques sur des fluctuations de la luminance dues au bruit. Un minimum de 6 moments géométriques est requis, deux ou trois dizaines permettent d'obtenir une précision suffisante pour cette phase de positionnement du robot appelée positionnement grossier comme expliqué plus loin. L'exploitation d'une combinaison de ces moments géométriques, par exemple caractérisant le centre de masse, l'axe principal d'inertie ou le rayon d'inertie de l'image traitée est généralement préférée à une utilisation brute de chacun d'entre eux. Ces moments géométriques sont également calculés sur l'image de référence lors des étapes 40 à 42, de la même manière que lors des étapes 37 à 39.

Ces moments géométriques, composantes du vecteur  $f$  et  $f_{ref}$  sont comparés termes à termes à l'étape 43 pour fournir les composantes du vecteur différence  $df$  à l'étape 44. Ce sont ces composantes qui alimentent le réseau de neurones.

Ces descripteurs globaux permettent de limiter considérablement le nombre d'informations à traiter. En contre partie, la précision de positionnement obtenue lors de l'exploitation de ces descripteurs est moins bonne et dépend essentiellement du nombre de descripteurs retenus, donc du nombre d'entrées du réseau de neurones.

Dans le but de limiter ce nombre qui peut devenir très important et donc conduire à des calculs complexes pour les précisions généralement souhaitées tout en conservant une bonne précision de positionnement, il est très avantageux d'associer le réseau de neurone précédemment décrit et recevant les descripteurs globaux à un nouveau

réseau de neurones recevant des informations provenant d'une corrélation d'images. Le positionnement se réalise alors en deux phases :

- une première phase appelée positionnement grossier et utilisant les descripteurs globaux pour un positionnement du robot au voisinage de la position de consigne c'est à dire selon une précision donnée,

- une deuxième phase appelée positionnement fin, qui entre en application après que le robot ait été rapproché de la position finale à la précision donnée près et qui met en oeuvre un deuxième réseau de neurones.

Cette deuxième phase est explicitée à l'aide de l'organigramme représenté à la figure 8.

L'étape 45 correspond à la saisie de l'image de référence définissant la position finale souhaitée du robot. Une étape 46 réalise un choix de fenêtres caractéristiques dans cette image de référence. Ce choix peut être réalisé manuellement par l'opérateur qui, par exemple, pointe un marqueur correspondant au centre de la fenêtre sur l'image de référence. Il peut également être effectué de manière automatique, par exemple à partir des paramètres définissant le motif lumineux et permettant de calculer la position de points particuliers ou mesurant l'entropie de l'image. Les positions de fenêtres sont ensuite mémorisées. Les paramètres définissant ces positions forment les composantes du vecteur  $f_{ref}$ . Les fenêtres correspondent préférentiellement à des points particuliers du motifs lumineux, car, comme il sera vu plus loin, elles permettent la mise en correspondance de portions de l'image courante avec des portions de l'image de référence et la superposition est d'autant plus précise que le motif inclu dans la fenêtre permet d'éviter les ambiguïtés de positionnement par superposition, c'est à dire corrélation. Ainsi un point est à préférer à une ligne , 2 lignes à 1 seule ligne, etc... Le nombre de fenêtres choisies est au moins de 3.

Le positionnement approximatif du robot ayant été effectué lors de la phase précédemment décrite de positionnement grossier, l'image courante visualisée par la caméra est saisie lors de l'étape 47. Elle est traitée à l'étape 48 de manière à ne prendre en compte que des zones de



recherche, de préférence plus grandes que les fenêtres caractéristiques, centrées sur ces fenêtres selon les informations mémorisées à l'étape 46.

L'étape suivante 49 effectue, dans chaque zone de recherche, une corrélation avec la fenêtre caractéristique provenant de l'étape 46 et  
5 correspondant à cette zone de recherche, par déplacement en x et y de cette fenêtre. Le pic de corrélation correspond à la superposition, pas forcément parfaite, des motifs. L'étape 50 calcule la position de la fenêtre caractéristique correspondant à ce pic de corrélation, c'est à dire les paramètres définissant sa position, par exemple les coordonnées du centre  
10 de la fenêtre. Ce calcul est effectué pour toutes les fenêtres, l'ensemble des paramètres représentant les composantes du vecteur  $f$ . Le déplacement en x et y de la fenêtre caractéristique est généralement suffisant pour effectuer la corrélation. Un prétraitement de l'image, lissage, seuillage....pourra être réalisé. Il est alors sous-entendu à l'étape 45 et 47.  
15 Le nombre de fenêtres doit être au moins de trois pour permettre de traiter les rotations d'image. Les vecteurs  $f$  et  $f_{ref}$  sont ensuite comparés termes à termes à l'étape 51 pour fournir les composantes du vecteur  $df$  à l'étape 52.

Le réseau de neurones appelé n° 2, alimenté par ces  
20 composantes, est différent de celui n°1 exploité par les descripteurs globaux. Ce réseau de neurones n° 2 a également subi une phase d'apprentissage. Les exemples sont des images très voisines du point de positionnement final du robot.

La base de données correspond, pour moitié des exemples, à la  
25 phase de positionnement grossier, pour l'autre moitié, à la phase de positionnement fin. Mais cette base peut être exploitée dans sa totalité par le réseau de neurones n° 1 lié au positionnement grossier et toute autre répartition est possible.

Les différentes opérations nécessaires au fonctionnement du  
30 système complet, robot, caméra, circuits de traitement algorithmique, réseaux de neurones, circuits de commande du robot sont décrites ci-après.

Le positionnement exact du robot, muni de sa caméra solidaire, à sa valeur de consigne est réalisé lors d'une phase préparatoire. Les  
35 circuits de traitement algorithmique calculent alors et mémorisent le vecteur

$f_{ref}$  correspondant au positionnement grossier, c'est à dire les moments géométriques de l'image de référence obtenus pour cette position, et le vecteur  $f_{ref}$  correspondant au positionnement fin, c'est à dire les coordonnées des centres des fenêtres caractéristiques choisies par l'opérateur. La position et les images des fenêtres caractéristiques de l'image de référence sont également mémorisés pour être transmis ultérieurement aux circuits de traitement algorithmique de l'image courante.

La phase d'apprentissage réalise le calcul des poids synaptiques du réseau de neurone n° 1, les sorties  $df$  étant alors reliées aux entrées de ce réseau de neurones. Les sorties du réseau, qui correspondent à une commande de déplacement du robot pour une image d'apprentissage, sont comparées aux valeurs de déplacement nécessaire du robot pour arriver à la position de consigne à partir de la position fournissant l'image d'apprentissage. Cette erreur est exploitée pour ajuster le poids des neurones, par exemple selon l'algorithme classique de rétropropagation. L'apprentissage de ce réseau terminé, les entrées du réseau de neurones n° 2 sont connectées aux sorties  $df$  pour un apprentissage similaire de ce deuxième réseau, les images d'apprentissage étant ici très voisines de l'image de référence. L'apprentissage des 2 réseaux terminé, le système est prêt à fonctionner. La phase active s'effectue alors en deux temps :

1er temps : les entrées du réseau de neurones n°1 sont alimentées par les composantes du vecteur  $df$  et les sorties de ce réseau sont connectées au circuit de commande du robot. Ces sorties fournissent les informations de déplacement qui positionnent le robot, en le rapprochant petit à petit de la position finale. Lorsque la valeur de ces sorties devient inférieure à un seuil fixé, le passage au deuxième temps s'effectue.

2ème temps : le réseau de neurones n°1 est déconnecté. Le réseau de neurones n° 2 est connecté, pour ses entrées, aux sorties des circuits de traitement algorithmiques fournissant les composantes des vecteurs  $df$ , par ses sorties aux circuits de commande du robot. La caméra est reliée à l'entrée des circuits de traitement algorithmique ci-dessus mentionnés et spécifiques à cette phase de positionnement fin, qui

calculent le vecteur  $d\vec{f}$ . Les sorties du réseau de neurone n°2 commandent le robot jusqu'à sa position finale.

Les applications de ce procédé de guidage et de positionnement d'un robot sont multiples. En robotique mobile, elles concernent par exemple la conduite d'engins en convoi groupé : le véhicule "robot" suiveur est asservi sur des points lumineux du véhicule le précédant, points intrinsèques au véhicule ou projetés par exemple par faisceau laser ou infrarouge à partir du véhicule suiveur.

Elles concernent tout type d'aide au positionnement tel que aide à l'atterrissage d'hélicoptère, aide à la navigation et au guidage d'engins en phase terminale etc...

En robotique industrielle, les applications sont les robots d'assemblage de pièces, de saisie, de positionnement de composants ou même d'inspection puisqu'il s'agit de comparer des motifs projetés sur des surfaces.

En médecine, les applications sont l'aide à la chirurgie assistée par ordinateur, il s'agit par exemple du "recalage" d'un patient par rapport à une base de données pré-opératoire c'est à dire par rapport à de précédentes mesures effectuées sur ce patient. Il s'agit également de guidage d'outils, d'asservissement de micro-mécanismes....

## REVENDEICATIONS

1 - Procédé de positionnement d'un robot (1) évoluant dans un espace en 3 dimensions dont la scène (4) est visualisée par un capteur  
5 d'image (2) solidaire du robot (1), caractérisé en ce qu'un motif lumineux est projeté sur la scène dans le champ du capteur, en ce que la forme projetée du motif est fonction de la position relative du robot par rapport à la scène et en ce que le positionnement est réalisé par comparaison, en  
10 entier ou en partie, de la forme de ce motif de l'image courante à la forme finale du motif de l'image de référence correspondant à la position recherchée, par l'intermédiaire d'un réseau de neurones recevant en entrées, des informations relatives à la forme projetée du motif acquise par le capteur.

15 2 - Procédé selon la revendication 1, caractérisé en ce que le motif lumineux (11,16) est un ensemble de taches lumineuses et en ce que les informations (22) transmises au réseau de neurones sont les différences, terme à terme, entre les coordonnées des barycentres de ces taches dans l'image de référence (19) et celles dans l'image courante (14).

20

3 - Procédé selon la revendication 1, caractérisé en ce que le motif lumineux (23,29) est une ligne fermée et en ce que les informations (36) transmises au réseau de neurones sont les différences, terme à terme, entre les descripteurs de Fourier choisis pour cette ligne dans l'image de  
25 référence (32) et leur équivalent (26) dans l'image courante.

4 - Procédé selon la revendication 1, caractérisé en ce que les informations (44) transmises au réseau de neurones sont les différences, terme à terme, entre les moments géométriques choisis pour le motif  
30 projeté dans l'image de référence (42) et leur équivalent (39) dans l'image courante.

5 - Procédé selon la revendication 1, caractérisé en ce que la comparaison est effectuée par corrélation de zones sélectionnées, ou  
35 fenêtres, (46) de l'image de référence (45) avec des zones de recherches (48) correspondantes dans l'image courante (47) en déplaçant ces fenêtres

de l'image de référence dans ces zones de recherche et en ce que l'écart mesuré(52) entre la position des fenêtres dans l'image de référence et la position (49) de ces mêmes fenêtres correspondant à un pic de corrélation du motif pour la zone de recherche concernée, dans l'image courante,

5 alimente les entrées du réseau de neurones.

6 - Procédé de positionnement d'un robot (1), caractérisé en ce qu'une première étape réalise un positionnement grossier du robot en exploitant des descripteurs globaux de l'image selon le procédé de la revendication 3 ou 4, les différences (36,44) alimentant un premier réseau

10 de neurones dont les sorties commandent le positionnement du robot, et en ce qu'un deuxième réseau de neurones est mis en oeuvre dans une deuxième étape pour un positionnement fin du robot, les sorties de ce réseau commandant alors le robot, cette deuxième étape exploitant le

15 procédé d'asservissement par corrélation selon le procédé de la revendication 5.

7 - Robot (1) équipé d'au moins un capteur d'images (2) caractérisé en ce qu'il comporte une source lumineuse (3) solidaire du capteur d'image pour projeter, sur la scène (4) observée par le capteur, un

20 motif lumineux, et un circuit de traitement pour réaliser l'asservissement du robot par comparaison du motif courant projeté sur la scène (5) au motif correspondant à une position de consigne par l'intermédiaire d'au moins un réseau de neurone.

25

8 - Robot selon la revendication 7, caractérisé en ce que la comparaison est effectuée sur les descripteurs de Fourier du motif lumineux.

30 9 - Robot selon la revendication 7, caractérisé en ce que la comparaison est effectuée sur les moments géométriques du motif lumineux.

10 - Robot selon la revendication 7, caractérisé en ce que la

35 comparaison est effectuée par corrélation d'au moins une zone du motif

définissant la position de consigne avec la zone correspondante du motif courant projeté.

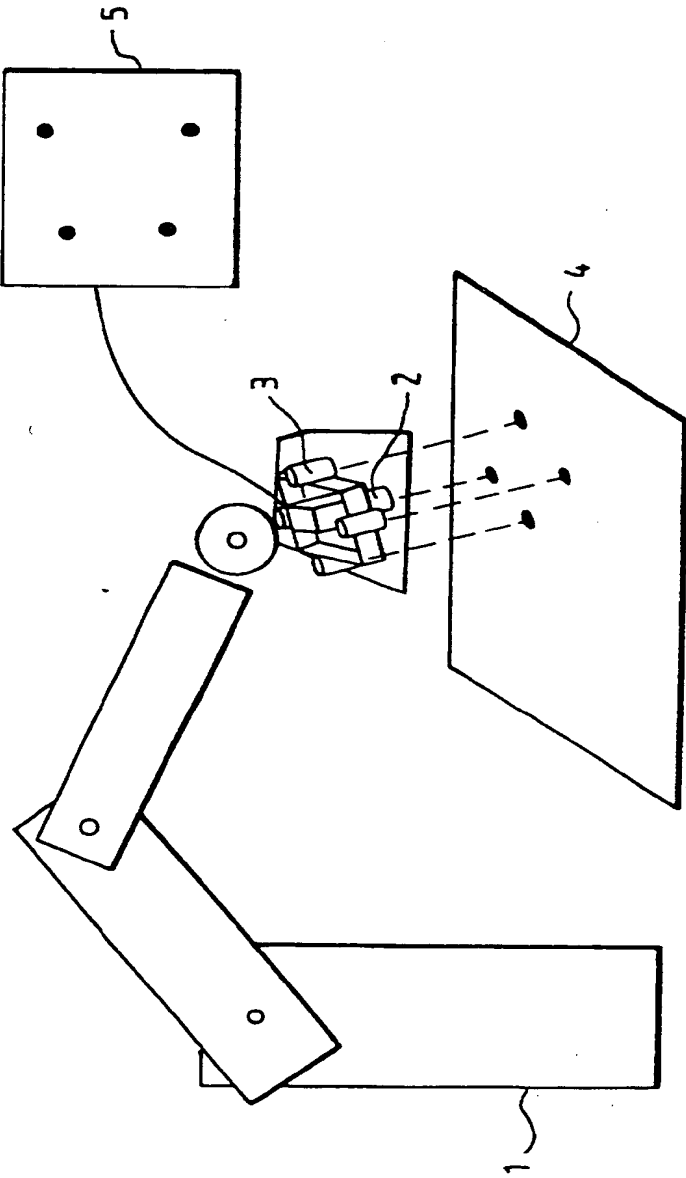


FIG.1

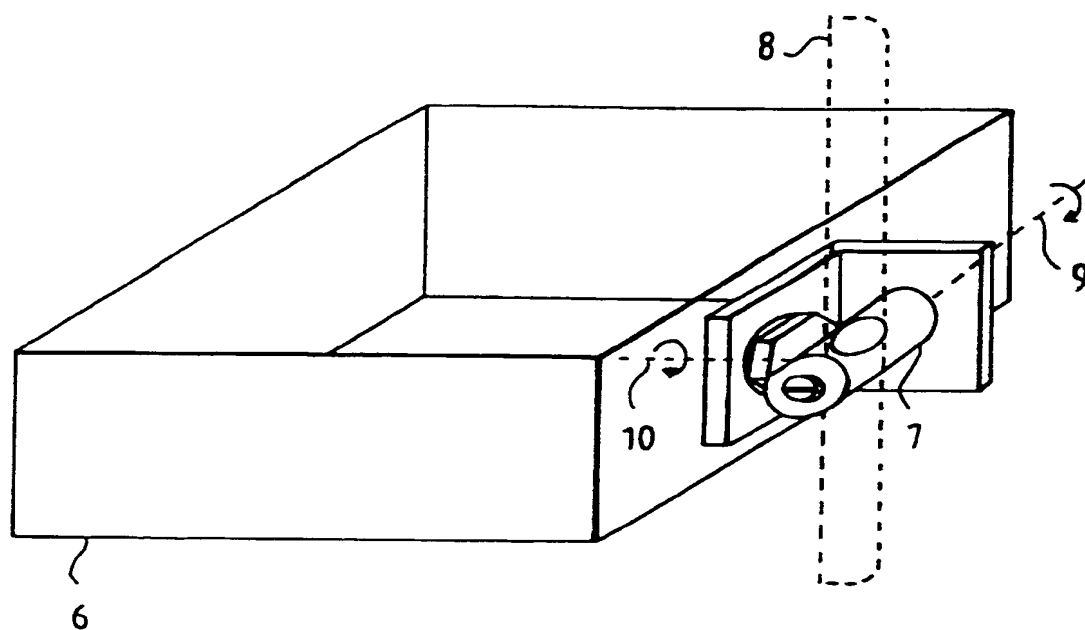


FIG.2



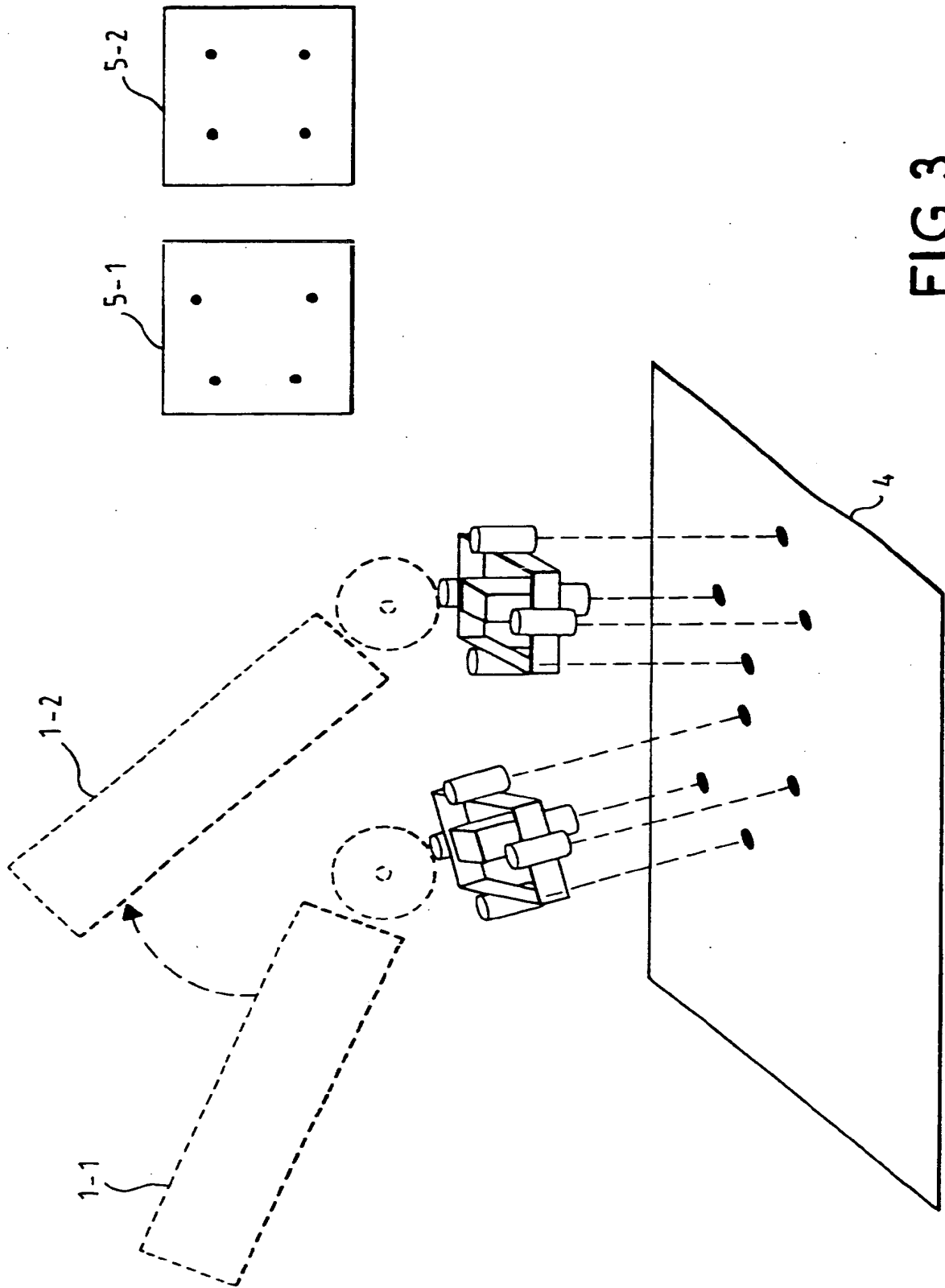


FIG. 3

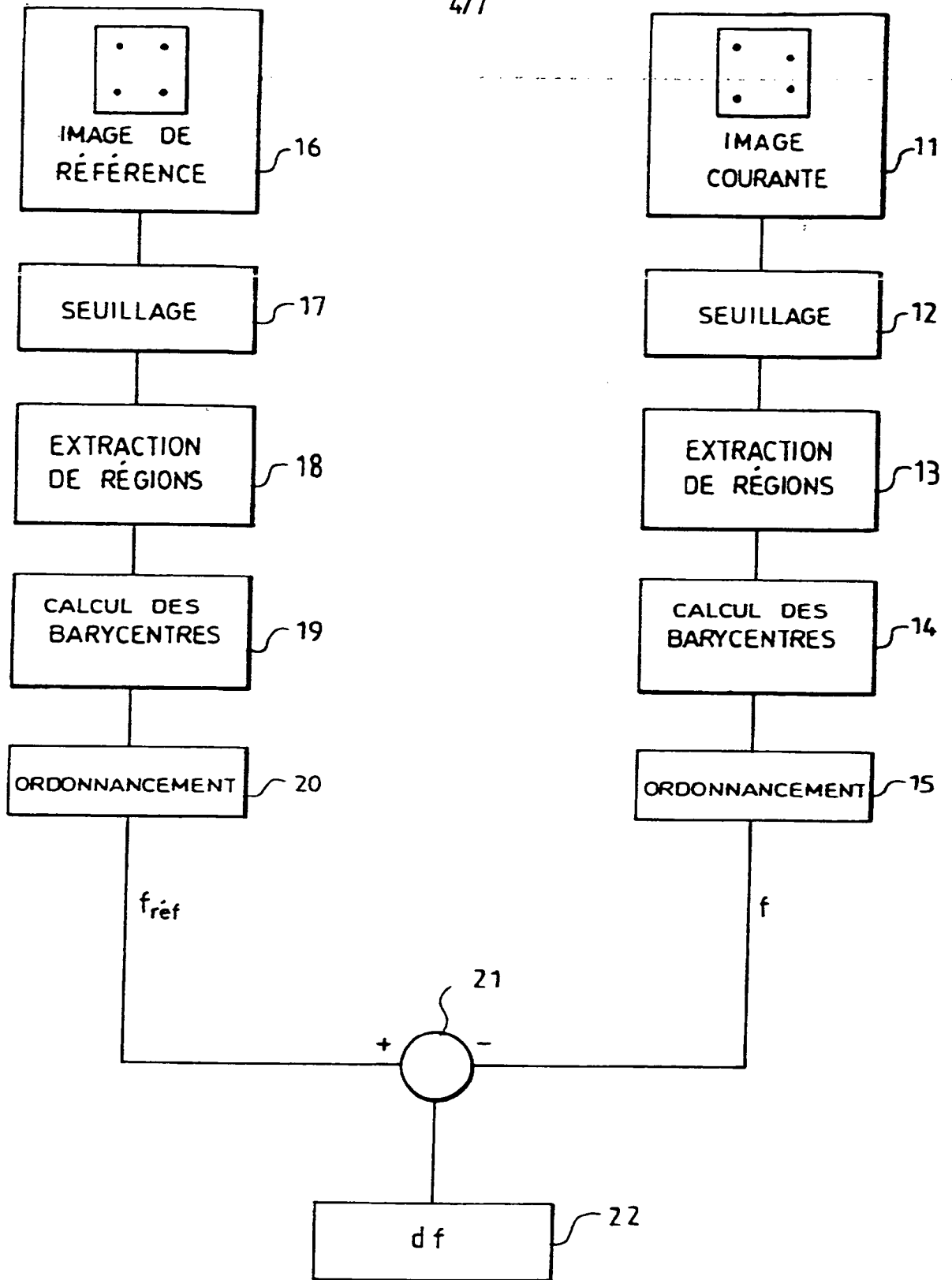


FIG.4

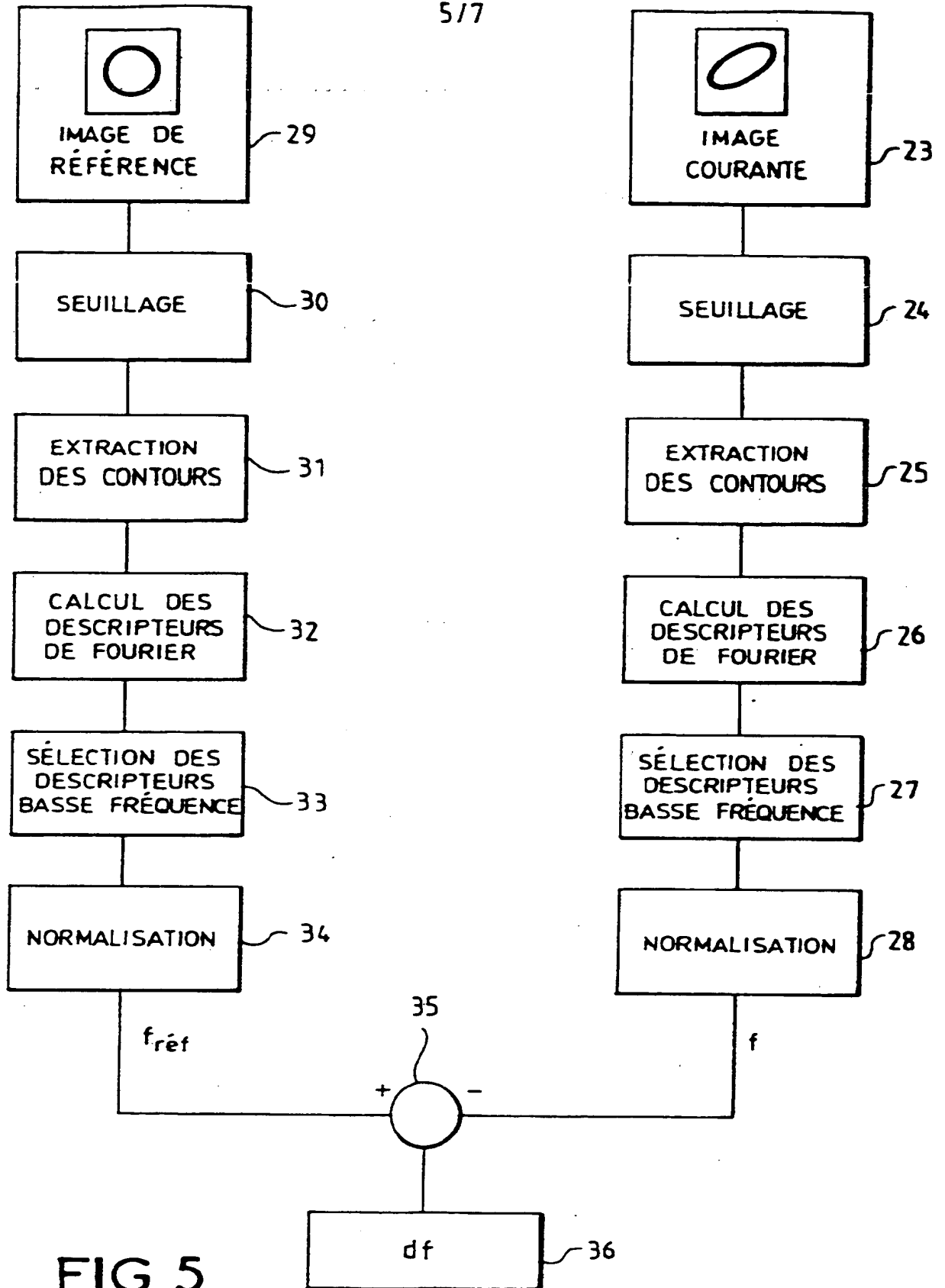
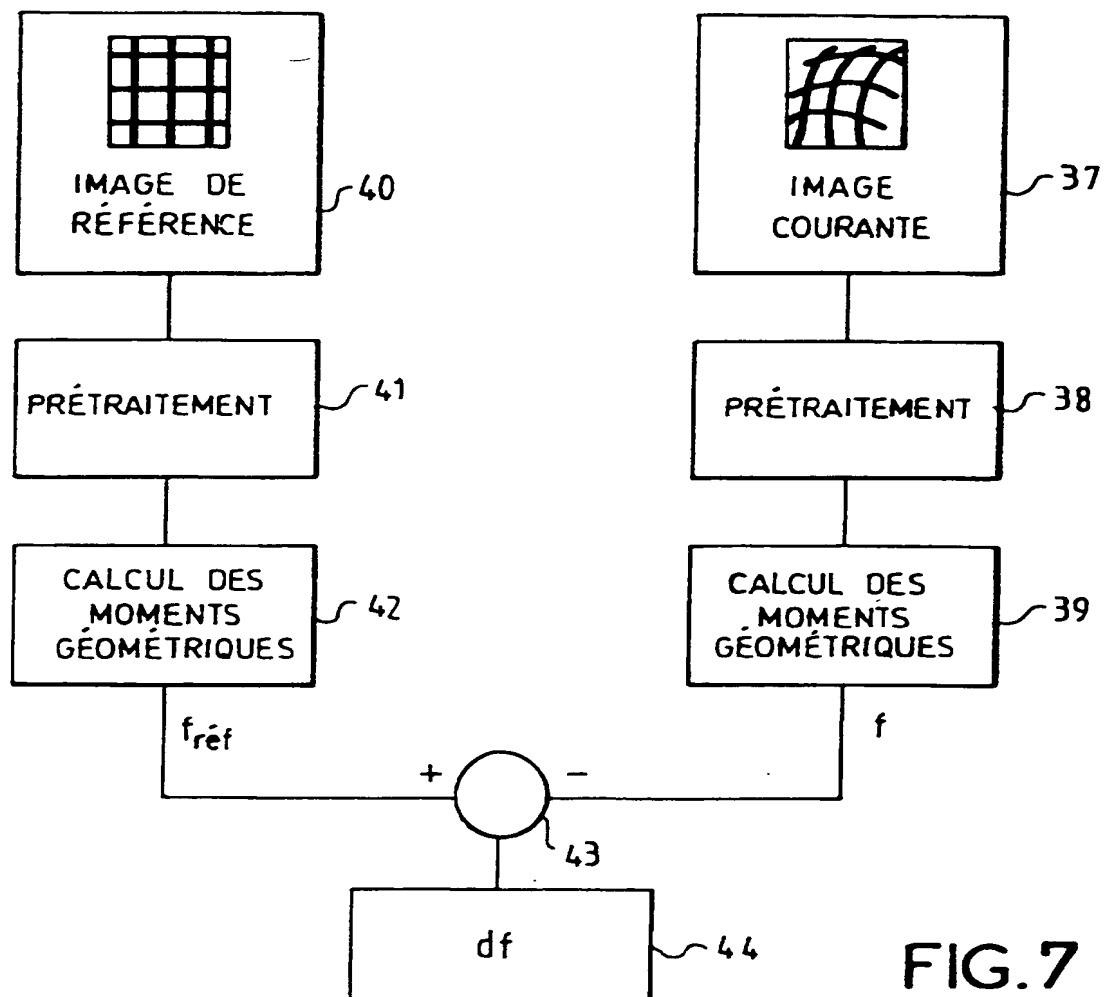
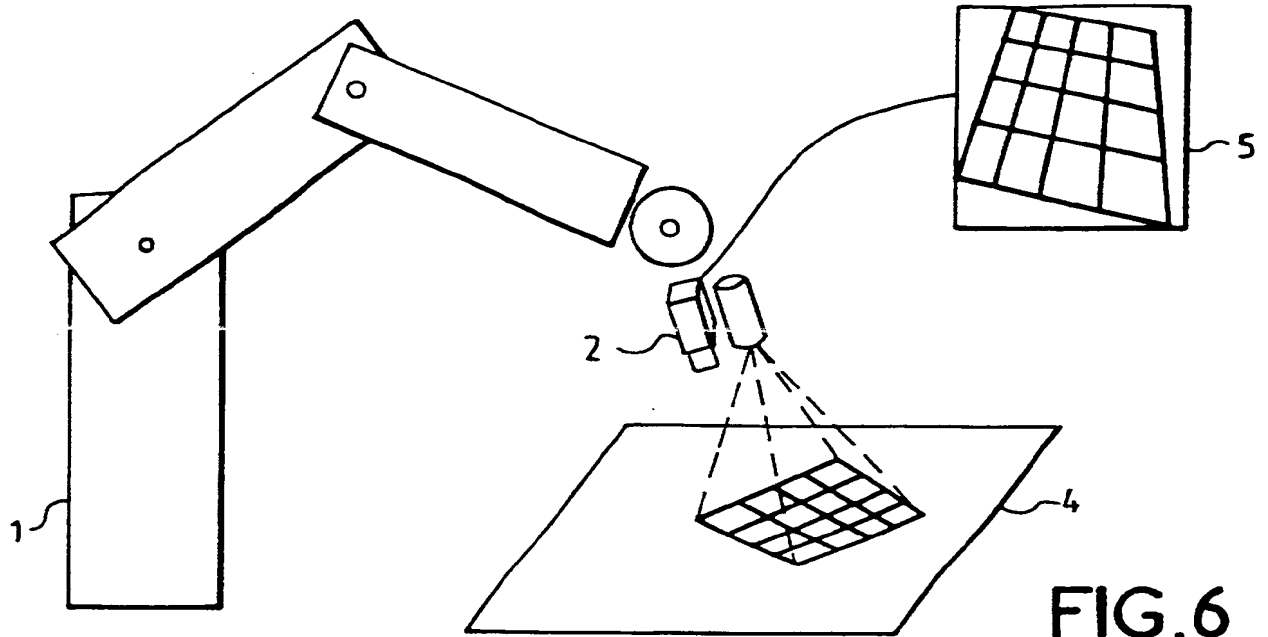


FIG.5

6/7



7/7

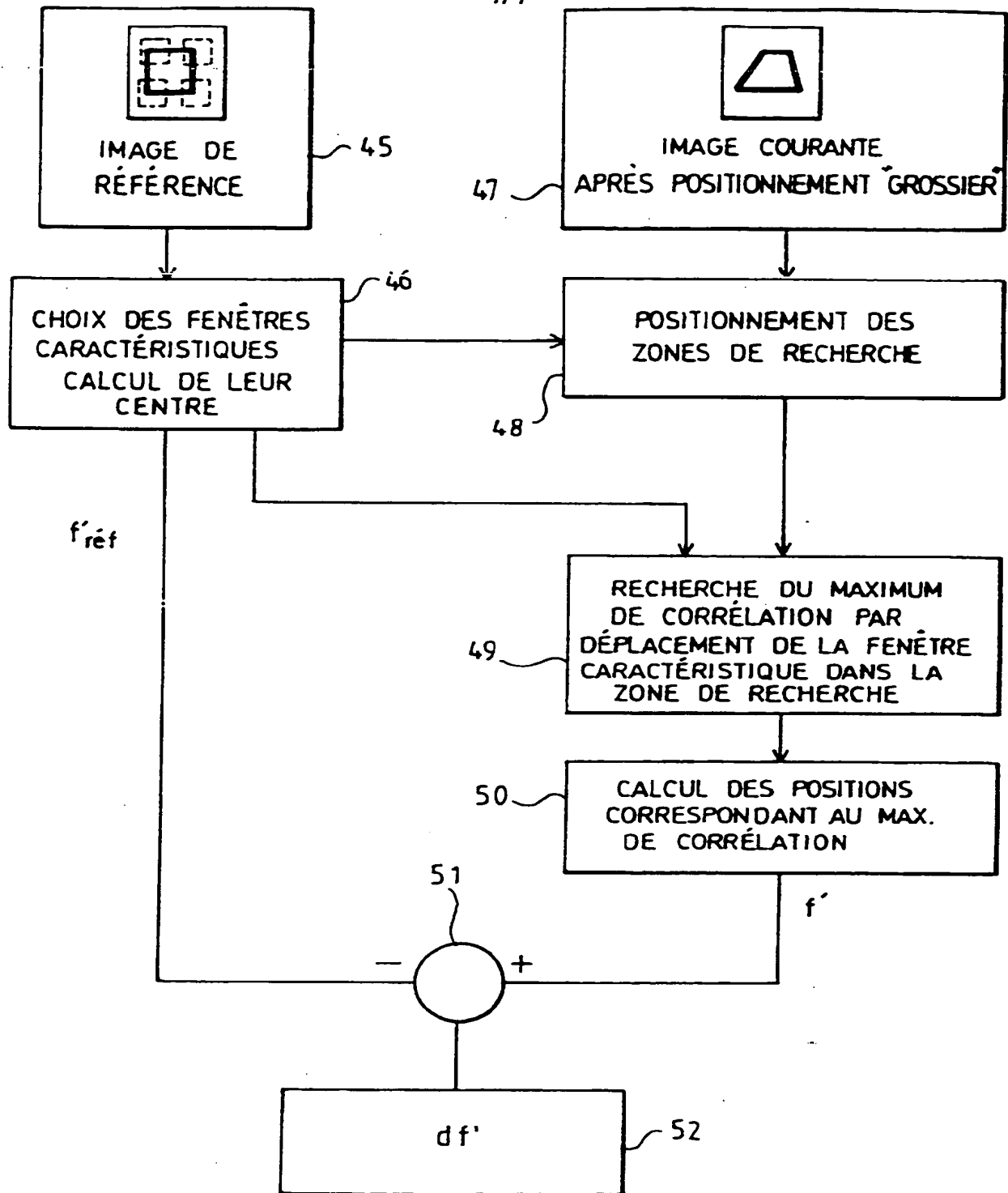


FIG.8

INSTITUT NATIONAL

RAPPORT DE RECHERCHE  
PRELIMINAIREN° d'enregistrement  
nationalde la  
PROPRIETE INDUSTRIELLEétabli sur la base des dernières revendications  
déposées avant le commencement de la rechercheFA 510980  
FR 9500117

DOCUMENTS CONSIDERES COMME PERTINENTS		Revendications concernées de la demande principale
Catégorie	Citation du document avec indication, en cas de besoin, des parties pertinentes	
X	IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS AND CONTROL INSTRUMENTATION, vol. 39, no. 6, Décembre 1992 NEW YORK US, pages 490-496, XP 000323856 HIDEKI HASHIMOTO ET AL. 'Visual Control of Robotic Manipulator based on Neural Networks' * page 490 - page 496 *	1,6
Y	---	2-5,7-10
Y	REVUE TECHNIQUE THOMSON-CSF, vol. 25, no. 1, Mars 1993 ISSN 0035-4279 , FRANCE, pages 107-125, HENOCQ H. ET BUREL G. 'Reconnaissance d'objets 3 D par analyse de Fourier de la silhouette' * page 107 - page 125 *	2,3,7,8
A	---	1,6
Y	ROBOTICS AND AUTONOMOUS SYSTEMS, vol. 11, no. 3/4, Décembre 1993 AMSTERDAM NL, pages 243-260, XP 000414331 J.R. COOPERSTOCK & E.E. MILIOS 'Self-supervised Learning for Docking and Target Reaching' * page 243 - page 260 *	4,9
Y	PATENT ABSTRACTS OF JAPAN vol. 6 no. 226 (E-141) ,11 Novembre 1982 & JP-A-57 129590 (MITSUBISHI D.K.K.) 11 Août 1952, * abrégé *	5,10
	---	
	-/-	
Date d'achèvement de la recherche		Examineur
12 Octobre 1995		Hauser, L
CATEGORIE DES DOCUMENTS CITES		
X : particulièrement pertinent à lui seul Y : particulièrement pertinent en combinaison avec un autre document de la même catégorie A : pertinent à l'encontre d'un motif une revendication ou schéma-plan technologique général O : divulgation non-écrite P : document intermédiaire  T : théorie en principe à la base de l'invention E : document de brevet bénéficiant d'une date antérieure à la date de dépôt et qui n'a été publié qu'à cette date de dépôt ou qu'à une date postérieure. D : cité dans la demande L : cité pour d'autres raisons  A : membre de la même famille, document correspondant		

2

EPO FORM 1500 (01.92) (P04C11)

INSTITUT NATIONAL  
de la  
PROPRIÉTÉ INDUSTRIELLE

RAPPORT DE RECHERCHE  
PRELIMINAIRE

établi sur la base des dernières revendications  
déposées avant le commencement de la recherche

FA 510980  
FR 9500117

DOCUMENTS CONSIDERES COMME PERTINENTS		Revendications concernées de la demande examinée
Catégorie	Citation du document avec indication, en cas de besoin, des parties pertinentes	
A	US-A-4 453 085 (TIMOTHY R. PRYOR) 5 Juin 1984 * colonne 4, ligne 35 - colonne 11, ligne 31; figures 1-8 *	1,5
A	EP-A-0 152 594 (LICENTIA PATENT-VERWALTUNGS-GMBH) 28 Août 1985 * figure *	1,6
A	EP-A-0 547 926 (AUTOMOBILES PEUGEOT ) 23 Juin 1993 * le document en entier *	1,6
A	US-A-4 884 216 (KUPERSTEIN M.) 28 Novembre 1989 * le document en entier *	1,6
A	REVUE TECHNIQUE THOMSON-CSF, FRANCE, vol. 25, no. 1, Mars 1993 ISSN 0035-4279, pages 65-82, HENOCQ H. ET VENAILLE C. 'Dynamic calibration of a CCD camera by inverse transformation'	1
		DOMAINES TECHNIQUES RECHERCHES (Int. CL. A)
Date d'achèvement de la recherche		Examineur
12 Octobre 1995		Hausser, L
<p><b>CATÉGORIE DES DOCUMENTS CITES</b></p> <p>X : particulièrement pertinent à lui seul Y : particulièrement pertinent en combinaison avec un autre document de la même catégorie A : pertinent à l'encontre d'au moins une revendication ou arrière-plan technologique général O : divulgation non-écrite P : document intercalaire</p> <p>T : théorie ou principe à la base de l'invention E : document de brevet bénéficiant d'une date antérieure à la date de dépôt et qui n'a été publié qu'à cette date de dépôt ou qu'à une date postérieure. D : cité dans la demande L : cité pour d'autres raisons &amp; : membre de la même famille, document correspondant</p>		

2

EPO FORM 150 (01.92) (P4C11)

**THIS PAGE BLANK (USPTO)**